UNIVERSITÄT SIEGEN

Fakultät IV
Department Mathematik

Probabilistic Forecasting of Medium-Term Electricity Demand: A Comparison of Time Series Models

Kevin Berk and Alfred Müller

University of Bristol

March 13, 2015

Load forecasting

Probabilistic forecasting of electricity demand is an important task for all active market participants!

- pricing of electricity contracts for end customers
- basis for risk management and hedging strategies of suppliers and vendors
- optimization of electricity procurement (generation) and consumption
- evaluation of the fair market price of electricity

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

Literature

- approaches vary from AI-based like neural networks to parametric approaches like exponential smoothing or time series analysis (Weron (2006), Alfares and Nazeeruddin (2002))
- typically considered: load forecasting models for households or for the total system load (Paatero and Lund (2005), Dordonnat et. al (2008))
- lack of information about uncertainty (deterministic forecasts)
- usual forecast horizon is hours to days ((very) short term)

Suitable studies of medium term probabilistic models for industrial end customers seem to be rare so far.

Our aim:

- load forecasting for companies with respect to business sectors
- medium-term, i.e. year-ahead forecasts
- total system load as an exogenous factor
- SARIMA model for stochastic load component
- challenge: consumption patterns can vary significantly among different sectors

Outline

1. Motivation

2. Load profiles

3. Methodology

4. Empirical analysis

5. Forecasting

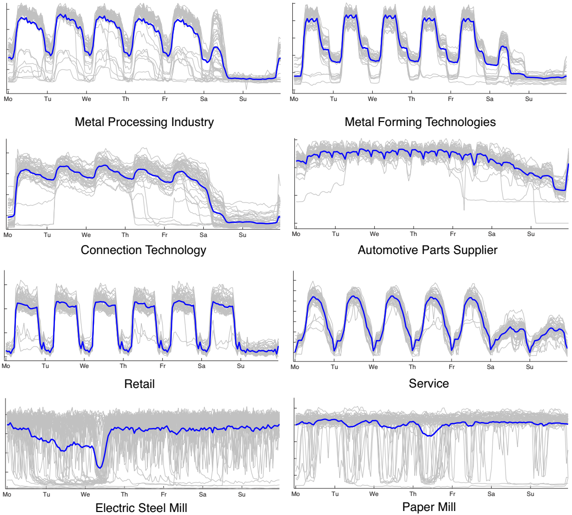
6. Comparing Models

7. Conclusion

Load profiles

- metering of electricity demand
- apply our model to hourly load data
- the cumulated load of a set of households or the total system load show a quite homogeneous behavior (diversification)
- in contrast, the load of a single industry customer is more heterogeneous
- reason: stochastic events like machine failures have a greater impact on the individual demand profile

Some examples



Fundamental model equation

Let C_t denote the electricity consumption at time t, the model equation has the following form:

$$\log C_t = D_t + u(G_t^*) + R_t$$

Here,

- D_t is a **deterministic trend**,
- $u(G_t^*)$ is a function of the **residual grid load** G_t^* and
- R_t is a **residual time series**.

A similar-day approach for D_t

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

The deterministic trend D_t is given by

$$\begin{aligned} D_t = &\beta_0 + \beta_1 \sin\{\frac{2\pi t}{365.25 \cdot 24}\} + \beta_2 \cos\{\frac{2\pi t}{365.25 \cdot 24}\} \\ &+ \sum_{j=3}^{16} \beta_j \vartheta_{j-2}(t) + \sum_{j=17}^{40} \beta_j \varrho_{j-16}(t), \end{aligned}$$

where the β_j are determined via OLS (using outlier-cleaned historical data). The dummies ϑ_j , $j = 1, \ldots, 14$ are described in table 1 and $\varrho_1, \ldots, \varrho_{24}$ denote the hours.

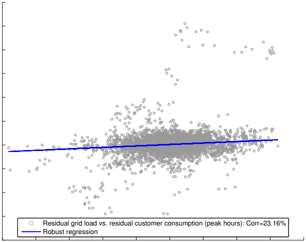
A similar-day approach for D_t

Table : Overview of dummy variables for D41sincos.

ϑ_1	Mondays (if not a public holiday or a bridge day)
ϑ_2	Fridays (if not a public holiday or a bridge day)
ϑ_3	Tuesdays, Wednesdays and Thursdays (if not a p. hol.)
ϑ_4	Saturdays (if not a public holiday)
ϑ_5	Sundays (if not a public holiday)
ϑ_6	Winter holiday period
ϑ_7	Summer holiday period
ϑ_8	Public holidays
ϑ_9	Bridge day (Monday before a public holiday)
ϑ_{10}	Bridge day (Friday after a public holiday)
ϑ_{11}	January 1st
ϑ_{12}	December 24th
ϑ_{13}	December 25th and 26th
ϑ_{14}	New Year's Eve

G_t^* as an exogenous factor

- G_t^* denotes the residual grid load, i.e. the total system load minus its deterministic part
- maps the stochastic fluctuations of the market
- correlation with a customer's electricity demand can be positive or negative
- the function $u(G_t^*)$ could be any suitable function, we found a linear approach sufficient



The residual time series R_t

- covers remaining autocorrelations and seasonalities
- often modeled by a Gaussian white noise process (iid normal distributed)
- general assumption of independence and light tails is wrong
- choose a (seasonal) ARIMA model for R_t
- number and choice of parameters depend on business sector

Database

Motivation

Load profiles

Methodology

Empirical analysis

Forecasting

Comparing Models

Conclusion

- load data of 21 German customers
- each data set includes an hourly profile of two years (17520 data points)
- hourly market and grid load data, provided by EEX and ENTSOE
- customers are divided into three industry sectors

1. Retail industry
2. Two shift operating industry
3. Three shift operating industry

Real consumption vs. D_t

Motivation

Load profiles

Methodology

Empirical analysis

Forecasting

Comparing Models

Conclusion

Deterministic trend / one year

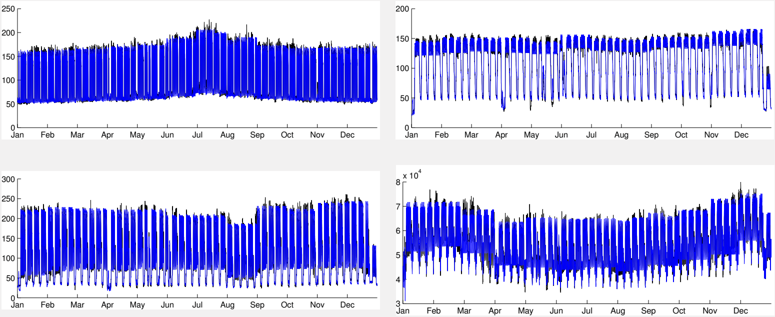


Figure : Retail industry (top left), three shift operating industry (top right), two shift operating industry (bottom left) and grid load (bottom right).

Real consumption vs. D_t

Deterministic trend / one week

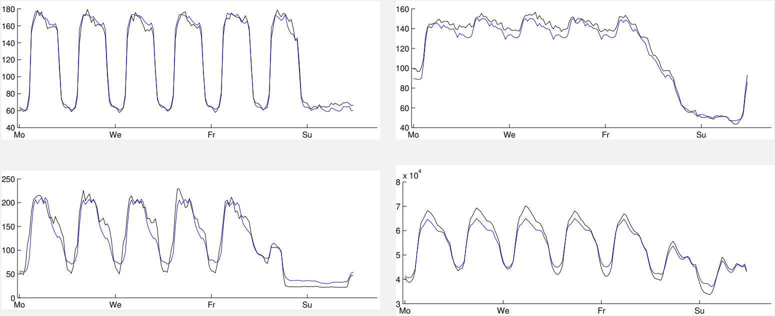


Figure : Retail industry (top left), three shift operating industry (top right), two shift operating industry (bottom left) and grid load (bottom right).

Function $u(G_t^*)$

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

$$u(G_t^*) = \lambda_{0,k} + \lambda_{1,k} G_{t,k}^* \quad \text{where } k \in \{\text{peak}, \text{off-peak}, \text{weekend}\}$$

- dependence between residual demand (stochastic load component) and residual system load
- linear approach for peak, off-peak and weekend hours
- negatively correlated for retail industry customers

Industry	$\rho_{\mathbf{p}}$	$\rho_{\mathbf{off}}$	$\rho_{\mathbf{we}}$	$\lambda_{0,\mathbf{p}}$	$\lambda_{1,\mathbf{p}}$	$\lambda_{0,\mathbf{off}}$	$\lambda_{1,\mathbf{off}}$	$\lambda_{0,\mathbf{we}}$	$\lambda_{1,\mathbf{we}}$
Retail	**-14.1 %**	**-18.4 %**	**-25.9 %**	16.18	-0.36	-309	-0.46	-420	-0.78
2 Shifts	**22.6 %**	**14.1 %**	**25.1 %**	552.15	0.68	931	0.6	295	2.31
3 Shifts	**23.1 %**	**31.0 %**	**24.3 %**	195.96	0.31	595	0.22	1109	0.55

Residual time series R_t

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

$$R_t = \log C_t - D_t - u(G_t^*)$$

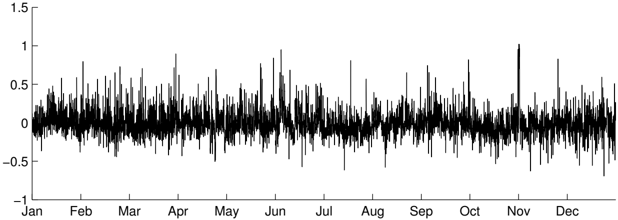


Figure : Residual time series R_t for retail industry

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

Model choice for R_t

- BIC suggests a SARIMA$(p,0,q)\times(P,0,Q)_{24}$ model with

$$(p,0,q)\times(P,0,Q)=$$

$$=\begin{cases}(3,0,3)\times(0,0,2) & \text{for retail customers,}\\ (4,0,4)\times(1,0,1) & \text{for two shift operating customers,}\\ (2,0,0)\times(2,0,2) & \text{for three shift operating customers.}\end{cases}$$

- innovations are heavy tailed

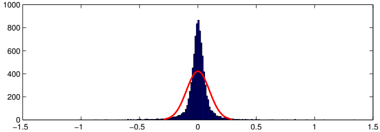


Figure : normal distribution fit for SARIMA model innovations

Model choice for R_t

Best fit for innovations given by NIG-distribution.

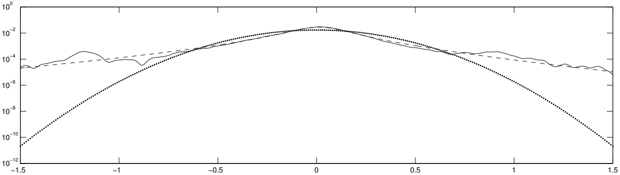
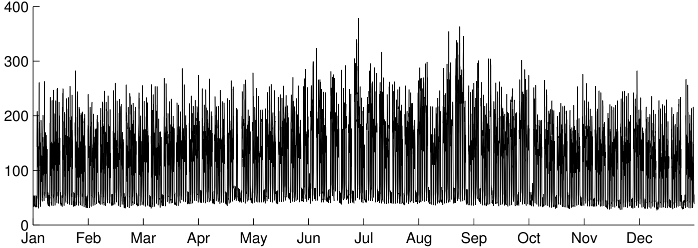


Figure : Fit of a normal distribution (dotted) and a normal inverse Gaussian distribution (dashed) to the kernel density of the empirical innovations (solid) on a semi-logaritmic scale.

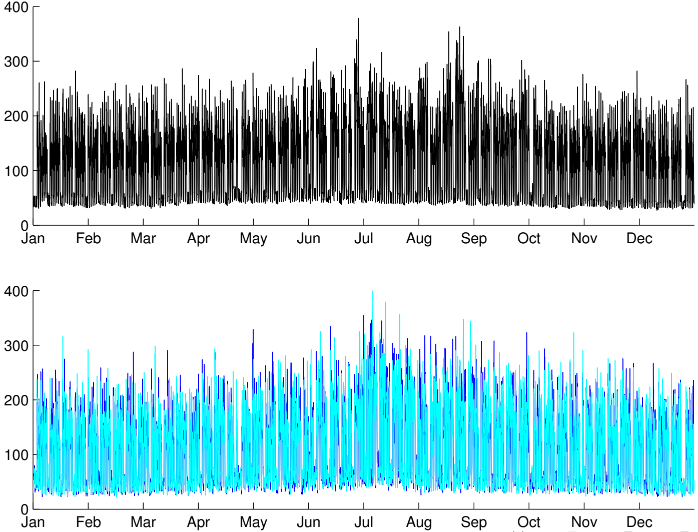
Medium-term forecasts

- year ahead forecast scenarios
- Monte-Carlo approach for retail pricing (see Burger and Müller (2012))
- trading strategy evaluation
- investment appraisal (thermal power station, photovoltaic system)
- peak load management
- strategic planning

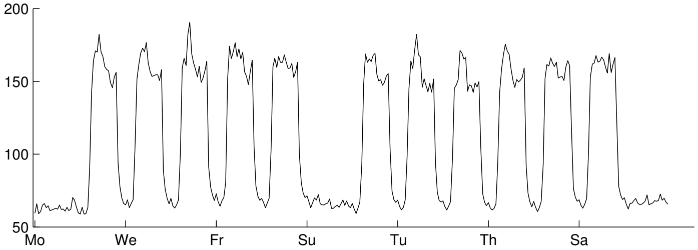
Real load vs one-year forecast scenarios



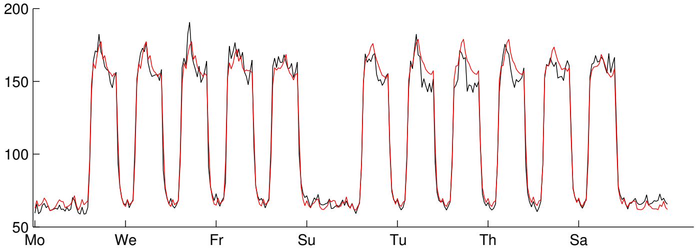
Real load vs one-year forecast scenarios



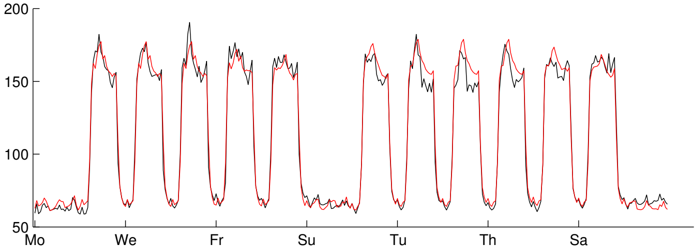
Real load vs expected consumption / forecast scenarios

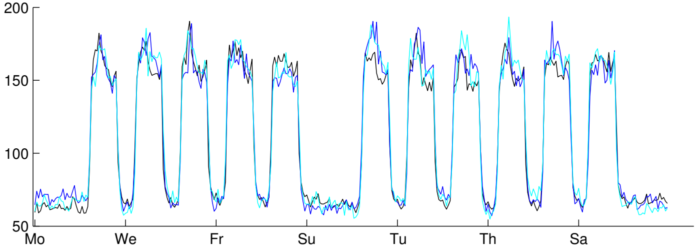


Real load vs expected consumption / forecast scenarios



Real load vs expected consumption / forecast scenarios





Comparing Models

Comparing different variants of the model:

- different number of parameters for seasonal patterns
- monthly dummies, sin-cos, area-preserving splines of order 4
- innovations normal, t or NIG

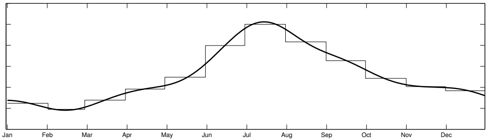


Figure : Regression coefficients of monthly dummy variables and fitted spline of order four.

Evaluation of probabilistic forecasts

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

For a given observation $\tilde{x}$ let $\tilde{F}(z) = 1_{[z \geq \tilde{x}]}$ be the empirical cdf. We use the *continuous ranked probability score* (CPRS), which is defined for a probabilistic forecast with cdf F and an observation $\tilde{x}$ as

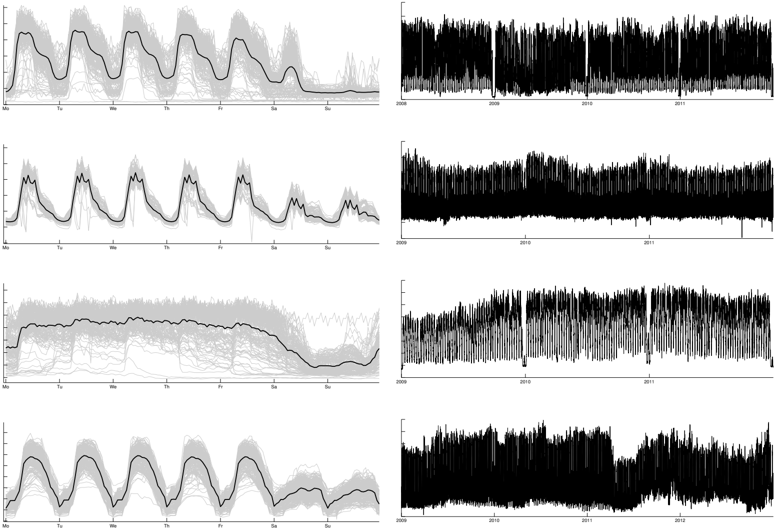
$$\begin{aligned}\mathrm{CRPS}(F, \tilde{x}) &= \int_{-\infty}^{\infty} (F(z) - \tilde{F}(z))^2 dz \\ &= \int_0^1 \mathrm{QS}_\alpha(F^{-1}(\alpha), \tilde{x}) d\alpha\end{aligned}$$

with

$$\mathrm{QS}_\alpha(q, \tilde{x}) = 2(1_{[\tilde{x} < q]} - \alpha)(q - \tilde{x})$$

the *quantile score*.
We estimate F via Monte-Carlo simulation.
We compute this for every point of time t and add up over t (of one year).

Data of four companies



Evaluation of probabilistic forecasts

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

Table : CRPS of different models for customer I to II. Absolute value and percentage deviation from benchmark. Rank in brackets, best values in bold.

Model	CRPS customer I	pct.	CRPS customer II	pct.
Benchmark	28.59 (10)	100%	14.24 (10)	100%
D31sincos-Normal	23.43 (9)	-18.03%	9.01 (9)	-36.77%
D35-SARIMA-Normal	23.26 (8)	-18.62%	8.95 (8)	-37.13%
D35-SARIMA-NIG	22.84 (7)	-20.12%	8.89 (5)	-37.6%
D35spline-SARIMA-NIG	22.78 (6)	-20.33%	8.93 (7)	-37.33%
D41sincos-SARIMA-NIG	21.29 (3)	-25.53%	8.83 (3)	-38.03%
D41sincos-GL-SARIMA-NIG	**20.57 (1)**	**-28.03%**	**8.72 (1)**	**-38.8%**
D51-SARIMA-Normal	21.96 (5)	-23.19%	8.89 (6)	-37.56%
D51-SARIMA-NIG	21.57 (4)	-24.55%	8.86 (4)	-37.78%
D51spline-GL-SARIMA-NIG	20.87 (2)	-27%	8.73 (2)	-38.68%

Evaluation of probabilistic forecasts

Table : CRPS of different models for customer III and IV. Absolute value and percentage deviation from benchmark. Rank in brackets, best values in bold.

Model	CRPS customer III	pct.	CRPS customer IV	pct.
Benchmark	888.78 (10)	100%	45.33 (10)	100%
D31sincos-Normal	830.72 (6)	-6.53%	27.46 (3)	-39.43%
D35-SARIMA-Normal	848.54 (9)	-4.53%	27.44 (2)	-39.47%
D35-SARIMA-NIG	844.13 (8)	-5.02%	27.59 (7)	-39.14%
D35spline-SARIMA-NIG	841.13 (7)	-5.36%	27.56 (6)	-39.21%
D41sincos-SARIMA-NIG	**771.95 (1)**	**-13.14%**	27.51 (5)	-39.31%
D41sincos-GL-SARIMA-NIG	774.24 (2)	-12.89%	27.97 (8)	-38.3%
D51-SARIMA-Normal	792.51 (3)	-10.83%	**27.39 (1)**	**-39.58%**
D51-SARIMA-NIG	792.53 (4)	-10.83%	27.49 (4)	-39.37%
D51spline-GL-SARIMA-NIG	795.26 (5)	-10.52%	28.14 (9)	-37.91%

Evaluation of probabilistic forecasts

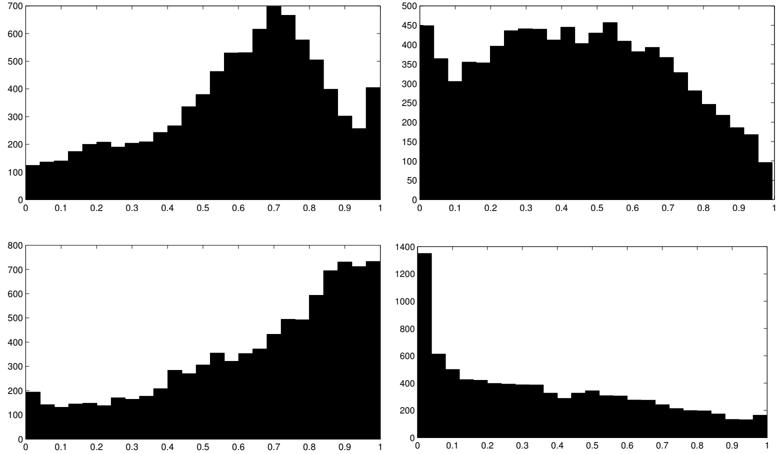


Figure : Rank histograms of the empirical PITs for the best ranked models. Top row: customer I and II. Bottom row: customer III and IV.

Conclusion

- model choice justified by data
- in-sample fit superior to other models (e.g. white noise residuals)
- innovations non-normal
- captures peak behavior well
- out-of sample backtesting supports the model
- good performance in risk management

Outlook

- Current research: Model extension to regime-switching load profiles
- ARMA model with regime dependent coefficients
- time-varying transition probabilities
- evaluating dependence modeling of probabilistic forecasts

Literature I

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

R. Weron.
Modeling and Forecasting Electricity Loads and Prices.
Chichester: Wiley, 2006.

K. Berk.
Modeling and Forecasting Electricity demand: A Risk Management Perspective
Springer Spektrum, 2015.

K. Berk and A. Müller.
Probabilistic Forecasting of Medium-Term Electricity Demand: A Comparison of Time Series Models.
submitted, 2015.

M. Burger and J. Müller.
Risk-adequate pricing of retail power contracts.
The Journal of Energy Markets, 4:53-75, 2012.

Literature II

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

M. Burger, B. Klar, A. Müller, G. Schindlmayr.
A spot market model for pricing derivatives in electricity markets.
Quantitative Finance, 4:109-122, 2004.

Motivation
Load profiles
Methodology
Empirical analysis
Forecasting
Comparing Models
Conclusion

Thank you for your attention!